Submitted to the AMIA Annual Symposium 2013.
draft material -- please do not cite

A mapping of RxNorm to the ATC/DDD Index helps analyze US prescription lists

Olivier Bodenreider, MD, PhD, Laritza M. Taft, MD, PhD
Lister Hill National Center for Biomedical Communications, National Library of Medicine, National Institutes of Health, Bethesda, MD

Abstract

Objectives*: To evaluate the suitability of the ATC/DDD Index (ATC) for analyzing prescription lists in the US.* ***Methods****: We mapped RxNorm clinical drugs to ATC through their ingredients and routes of administration. We used this mapping to classify a large set of prescription drugs with ATC and compared the median daily dose prescribed for a given drug to the defined daily dose (DDD) in ATC.* ***Results****: Of the 11k single-ingredient drugs from the prescribable subset of RxNorm, 7655 (68%) mapped to ATC through both their ingredient and route of administration, representing 1693 (44%) of the ATC 5th-level codes for single-ingredient drugs. An ATC class could be assigned to 97% of the 87k prescription drugs we investigated. The median daily dose prescribed is within +/-50% of the DDD for 60% of the 522 ingredient-route pairs in the prescription list analyzed.* ***Conclusions****: Although the mapping of RxNorm ingredients to ATC is largely incomplete, the most frequently prescribed drugs in the prescription set we analyzed were covered.*

Introduction

The analysis of prescription data requires that the prescription dataset, the drug terminology to which the prescription drugs are coded, and the drug classification system used for analysis purposes be interoperable. In the US, the standard drug vocabulary is RxNorm. However, RxNorm does not provide information about pharmacologic classes or daily doses. As a consequence, additional drug information sources need to be used for the analysis of prescription lists. One such resource is the ATC/DDD Index (ATC). Unlike many other drug information sources, ATC is not integrated in RxNorm, and its compatibility with RxNorm needs to be assessed. Moreover, ATC is not primarily a clinical resource, and its suitability for the analysis of prescription data also requires assessment.

The objective of the present study is to evaluate the feasibility of using ATC to analyze prescription lists in the US. More specifically, we propose to create a mapping between clinical drugs in RxNorm and ATC drugs, taking into account not only the active moiety, but also the route of administration. As an application, we propose to leverage this mapping for the analysis of a prescription list in the US, from the perspective of characterizing the drugs prescribed with ATC classes, and to compare the doses prescribed for these drugs to the defined daily dose in ATC.

Background

This investigation leverages ATC, RxNorm and a large prescription list obtained from SureScripts.

ATC

The Anatomical Therapeutic Chemical (ATC) classification [http://www.whocc.no/atc_ddd_index/] is a system developed by the World Health Organization (WHO) is recommended for worldwide use to evaluate drug utilization statistics. The system includes drug classifications at 5 levels; anatomical, therapeutic, pharmacological, pharmaceutics and drugs or ingredients. Also included are defined daily doses and administration routes assigned to each drug in accordance to the therapeutic and pharmacological groups.

For example, drugs from the digitalis glycosides group are included in the anatomical group “Cardiovascular system”. The route of administration and the defined daily dose are listed for each of the four 5th-level drugs.

C CARDIOVASCULAR SYSTEM

C01 CARDIAC THERAPY

C01A CARDIAC GLYCOSIDES

C01AA Digitalis glycosides

ATC code	Name	DDD	U	Adm.R	Note

C01AA02	acetyldigoxin	0.5	mg	O
C01AA03	digitalis leaves	0.1	g	O
C01AA04	digitoxin	0.1	mg	O

Multiple publications outside the US use ATC to investigate drug interactions, adverse events, and prescription drug costs among others. In interest of validity of the studies utilizing the ATC system, the documentation recommends each country to assign a national center responsible for assigning ATC codes to products marketed in within its territory.

As previously noted, the ATC classification includes five levels; the first four levels group the drugs by organ or system on which they act and by their mechanism of action, depending on therapeutic, pharmacological and chemical properties. The fifth level is the chemical substance or ingredient(s). The nomenclature uses international nonproprietary names (INN). If an INN name is not available for a given substance, the United States Adopted Name (USAN) or British Approved Name (BAN) are chosen. The active ingredients in the classification include a wide range of chemical entities used in a variety of countries. New ingredients are not included in the ATC system until it is approved for pharmaceutical use by at least in one country. Only herbal medicinal products approved by regulatory authorities are included in the classification.

The chemical substances are classified according to the main therapeutic use of the main ingredient and assigned one ATC code, per route of administration, and DDD. Similar ingredients combinations with similar strengths have different ATC codes, one per each category. Slow release tablets of same ingredients share the same ATC code. Single ingredient drugs, and combined ingredients are assigned individual codes each.

The DDD is the assumed average maintenance dose per day for a drug used for its main indication in adults. The DDD is not necessarily the Prescribed Daily Dose; the latter depends on individual patient characteristics such as age and weight, and pharmacokinetic considerations.

The DDD is calculated based on adult weight of 70 lb. The DDD can be an average of doses from different countries and might even be a dose that is never prescribed but is the the reflection of the more commonly used strengths. A DDD based on average is also assigned for depot formulations and intermittent dosing. Topical products, sera, vaccines, antineoplastic agents, allergen extracts, anesthetics and contrast media are not assigned a DDD.

The 2013 edition of ATC used in this study contains 4516 5^{th}-level ATC drugs, of which 3863 correspond to single-active moieties (as opposed to combinations).

RxNorm

RxNorm is a standardized nomenclature for clinical drugs compiled from 11 drug source vocabularies, and maintained by the National Library of Medicine (NLM). A clinical drug is defined as a pharmaceutical product with therapeutic or diagnostic properties available to patients. A clinical drug includes the ingredient(s), strength or concentration, and form appropriate for the intended administration route (e.g., *Thyroglobulin 32 MG Oral Tablet*).

A drug form group (DFG) is a term type that groups the drug forms (DF). Drug forms in turn are administration vehicles such as pills, tablets, syringes and lotions. They correspond to drug form groups classified by administration route, oral, parental, and topical, etc.

An ingredient (IN) is the chemical component of the drug, and the strength is the quantity of the ingredient represented in units of measure for mass, volume, or concentration (e.g., *5 mL solution, 1mg/mL*).

Single ingredient drugs have a unique chemical component, and multiple ingredient drugs have more than one active chemical component (e.g., *Amoxicillin 250 MG / Clavulanate 125 MG Oral Tablet*).

The RxNorm API is a web service provided by the National Library of Medicine for access to RxNorm with via SOAP or REST. Functions are available to find the status of a given RxNorm code, and to navigate between different types of drug entities (e.g., clinical drug to dose form group).

SureScripts dataset

The prescription drug list is an unidentified list comprised of 100,000 clinical drugs dispensed to emergency room patients over a period of three months at Suburban Hospital in Bethesda, Maryland. Each drug includes an anonymized prescription identifier, clinical drug name, drug form, strength, prescribed amount, and the intake duration.

Related work

ATC has been used in a variety of clinical and other research applications, but rarely in the US.

Epidemiological studies to investigate drug use, and drug use compliance among different populations have utilized ATC for drug classification purposes [1]. ATC/DDD as a classification system was also used for surveillance of antibiotic treatments and compliance assessment, by comparison of prescribed drugs and doses to those recommended in ATC across different institutions and within the same institution across departments [2]. When evaluating pharmacy stock of broad-spectrum antibiotics using DDD in ATC as measurement units and pharmacy sales figures, ATC allowed the analysis of the most commonly used antibiotic classes in each ward per clinical specialty. The study concludes that observation periods longer than one month are required to establish real reliable trends [3]. In the evaluation of elderly people's knowledge of the purpose of their medications, the authors used the anatomical and pharmacological action classes to classify the questions and answers [4]. ATC has also been used to support the detection of adverse events in the EU-ADR project [5].

Different groups have used ATC chemical interactions and similarities to predict drug group membership to pharmacological and physiological levels [6]. ATC has been leveraged in research to establish automatic classification of missing ingredients in ATC using information extraction (IE) and machine learning (ML) techniques to extract knowledge from free text. The extracted knowledge was used to express drug chemical and systemic effects from free text and assign ATC class membership [7]. Work by one of the authors investigated the ATC based on clinical drug properties using the annotations in NDF-RT. The study demonstrated that the ATC is fairly homogenous on mechanism of action and physiologic effects, however only half of the ingredients in the ATC could be analyzed with the approach due to missing annotations in the NDF-RT [8].

The specific contribution of our work is to assess the usability of ATC with prescription data in the US, and in relation to RxNorm, the standard vocabulary for drugs in the US (in particular in the context of the Meaningful Use criteria).

Methods

In ATC, a defined daily dose (DDD) is assigned not to an ingredient, but to an ingredient with a specific route of administration. For example, the DDD for *acetylsalicylic acid* is 3 g for oral forms, but 1 g when administered parenterally. As a consequence, the mapping of RxNorm clinical drugs to ATC requires a match for both the ingredient and the route of administration. Our approach to mapping RxNorm clinical drugs to ATC can be decomposed in two steps: 1) mapping RxNorm ingredients to ATC ingredients, and 2) mapping routes of administration between RxNorm and ATC.

In this investigation, we restrict the scope of the mapping to single-ingredient clinical drugs in RxNorm and ATC, because combination drugs are often underspecified in ATC. For example, the ATC 5^{th}-level code *N05BC51* corresponds to *meprobamate, combinations*, and is distinct from the single-ingredient category for *meprobamate* (*N05BC01*), but without specifying which ingredients can be associated with *meprobamate* or what the DDD for *meprobamate* is in this case. Moreover, since our goal is to analyze prescription lists, we restrict the mapping to clinical drugs from the prescribable subset of RxNorm.

Mapping ingredients between RxNorm and ATC

In previous work, we investigated the mapping of ingredients between RxNorm and ATC [9]. Since the scope of the two resources is slightly different, the mapping is not expected to be complete. For example, RxNorm includes several hundred allergenic extracts (e.g., *papaya allergenic extract 50 MG/ML Injectable Solution*) that are out of the scope of ATC. Conversely, diagnostic and therapeutic radiopharmaceuticals (e.g., *technetium* (^{99m}Tc) *bicisate*) are present in ATC (under *V09* and *V10*), but out of the scope of the prescribable subset of RxNorm.

We use the RxNorm API [http://rxnav.nlm.nih.gov/] to perform exact (case-insensitive) and normalized matches automatically against RxNorm ingredients for all the 5^{th}-level drugs in ATC. For example, the ATC term *meprobamate* (*N05BC01*) is an exact match for the RxNorm ingredient *Meprobamate* (*6760*) and the ATC term *ferric sodium gluconate complex* (*B03AC07*) is a normalized match for the RxNorm ingredient *Sodium ferric gluconate complex* (*261435*). The accuracy of the automatic mapping through normalization has been evaluated in previous work and deemed sufficient for research applications [9].

In a small number of cases, we use more aggressive normalizations techniques in order to account for the specific forms encountered in ATC, such as the mention of alternative names in parentheses in the drug term. For example the term for the ATC code *C10AD05* is *nicotinyl alcohol (pyridylcarbinol)*. We systematically remove such trailing parenthetical expressions and map the rest of the drug name to RxNorm, here, to *Nicotinyl Alcohol* (*7414*).

In most cases, only active moieties are represented in ATC, whereas RxNorm also covers their various salts, esters and complexes (e.g., *amoxicillin* vs. *amoxicillin trihydrate*). When a salt is present in both ATC and RxNorm, the mapping is made at this level (e.g., *levothyroxine sodium*). In most cases, however, RxNorm salt ingredients are mapped to the corresponding active moiety in ATC (e.g., *amoxicillin trihydrate* to *amoxicillin*).

In contrast to RxNorm, in which each ingredient is represented only once, ATC can have multiple codes for the same active moiety, depending on the anatomical system or therapeutic domain in which it is used. As a consequence, there will often be multiple ATC mappings for a given RxNorm ingredient. For example, the RxNorm ingredient *Ketoconazole* (*6135*) maps to the following ATC codes for this drug: *D01AC08* (from the *ANTIFUNGALS FOR DERMATOLOGICAL USE*), *G01AF11* (from the *GYNECOLOGICAL ANTIINFECTIVES AND ANTISEPTICS*) and *J02AB02* (from the *ANTIMYCOTICS FOR SYSTEMIC USE*).

Mapping routes of administration between RxNorm and ATC

In order to map routes of administration between RxNorm and ATC, we harmonized the dose form groups in RxNorm and the administration codes in ATC. We also assigned administration codes to ATC drugs when they were missing.

Harmonization of administration codes between RxNorm and ATC. RxNorm and ATC have different ways of representing routes of administration. In RxNorm, the route is expressed through the dose form group (DFG), but RxNorm DFGs actually represent the dose form (e.g., *Pill*), the route (e.g., *Ophthalmic Product*) or a mix of both (e.g., *Oral Gel Product*). Many clinical drugs are associated with multiple DFGs, typically one for the dose form and one for the route. For example, *Ketoconazole 200 MG Oral Tablet* (*197853*) is associated with both *Pill* and *Oral Product*. Of the 45 DFGs in RxNorm, 22 represent dose forms exclusively, but some of them are indicative of topical products nonetheless (e.g. *Shampoo Product*).

ATC assigns administration codes to the drugs in scope for the defined daily dose (e.g., *O* for oral, *N* for nasal, etc.). In addition to the 22 administration codes, ATC defines 10 coarser administration routes. Although ATC does not provide a correspondence between administration codes and administration routes, this correspondence is usually trivial to establish. Missing from the list of ATC administration routes are entries for the routes of ophthalmologic, otic, stomatologic and other topical products, for which ATC typically does not provide DDDs.

We extended the list of 10 administration routes from ATC with *ophthalmologic*, *otic*, *stomatologic* and *topical*, adding *urethral*, as it exists as an administration code. We mapped all relevant DFGs from RxNorm to the extended list of 15 administration routes derived from ATC.

Assignment of missing administration codes in ATC. As mentioned earlier, one issue for mapping drugs between RxNorm and ATC is that ATC assigns administration codes only to a subset of its drug entities, as required for the DDD. In practice, drugs or which no DDD is asserted are also missing an administration code. These drugs typically include topical products and systemic drugs for which there are large inter-individual variations (sera, vaccines, antineoplastic agents, allergen extracts, general and local anesthetics and contrast media).

For these drugs, we used various strategies to automatically infer an administration code when it was missing. These techniques were applied in the following order.

1. The authors assigned one or more administration code manually, not to individual drugs, but to specific ATC groups (at various levels), based on clinical knowledge (e.g., the group *Enemas* (*A06AG*) was associated with the administration route *rectal*). The list of ATC groups is shown in Table 1, along with the corresponding administration codes assigned.
2. The authors assigned one or more administration code manually to specific expressions found in the labels of ATC groups (e.g., ATC drugs from groups, whose label contain the expression systemic were associated with the administration routes *oral* and *parenteral*). The list of expressions is shown in Table 2, along with the corresponding administration codes assigned.

3. All administration codes found in the drugs for a given 4th-level ATC group were transferred to the drugs with missing administration codes in the same group (e.g., the drug *alogliptin* (*A10BH04*) "inherits" the administration code *oral* from the other drugs in the group *Dipeptidyl peptidase 4 (DPP-4) inhibitors* (*A10BH*), namely *sitagliptin*, *vildagliptin*, *saxagliptin*, and *linagliptin*).

4. The administration code *oral* was assigned by default to any digestive drug (e.g., the digestive drug *tilactase* (*A09AA04*) was assigned the administration code *oral* by default).

5. The administration code *topical* was assigned by default to any drug that has not been assigned one in the previous steps (e.g., the drug *tetracycline* (*D06AA04*) was assigned the administration code *topical* by default).

Table 1. Administration codes assigned to specific ATC groups. (Terms in all caps denote high-level classes)

ATC group code	**ATC group label**	**Administration route**
J06	IMMUNE SERA AND IMMUNOGLOBULINS	*parenteral*
J07	VACCINES	*oral, parenteral*
L	ANTINEOPLASTIC AND	*oral, parenteral, topical*
V01	ALLERGENS	*oral, parenteral*
N01A	ANESTHETICS, GENERAL	*oral, parenteral, inhalation*
N01B	ANESTHETICS, LOCAL	*topical, parenteral, stomatologic*
V08	CONTRAST MEDIA	*parenteral*
V03AN	Medical gases	*inhalation*
B05	BLOOD SUBSTITUTES AND PERFUSION	*parenteral, topical*
S01	OPHTHALMOLOGICALS	*ophthalmic*
S02	OTOLOGICALS	*otic*
S03	OPHTHALMOLOGICAL AND OTOLOGICAL	*ophthalmic, otic*
A07EA	Corticosteroids acting locally	*rectal*
A06AG	Enemas	*rectal*
A01	STOMATOLOGICAL PREPARATIONS	*stomatologic*
R02	THROAT PREPARATIONS	*stomatologic*

Table 2. Administration codes assigned to specific expressions in the labels of ATC groups.

Expression in ATC group label	**Administration route**
oral	*oral*
nasal preparation	*nasal*
enema	*rectal*
vaginal	*vaginal*
inhalant	*inhalation*
dermatological preparation	*topical*
i.v.	*parenteral*
parenteral preparation	*parenteral*
systemic	*oral, parenteral*

Mapping RxNorm clinical drugs to ATC

Having mapped RxNorm ingredients to ATC and harmonized routes of administration between DFGs in RxNorm and (augmented) administration codes in ATC through the extended list of administration routes, we define a mapping between a clinical drug in RxNorm and an ATC 5th-level drug when the following two conditions are met.

1. The ingredient (active moiety or salt ingredient) of the clinical drug in RxNorm maps a 5th-level drug in ATC as defined above.

2. The dose form group for the clinical drug in RxNorm and the administration code (or one of the administration codes, if multiple) of the 5th-level drug in ATC are compatible (i.e., are associated through the same administration route), as defined above.

For example, the RxNorm clinical drug *Amoxicillin 25 MG/ML Oral Suspension* (*313797*) maps to the ATC code *J01CA04* (*amoxicillin*), because the ingredient of the RxNorm clinical drug, *amoxicillin*, maps to this ATC code, and the dose form group of the RxNorm clinical drug, *Oral Product*, matches one of the routes of administration for the ATC code *J01CA04*, *O*, through the administration route *oral* (See Table 3). In contrast, despite the fact that both drugs have the same ingredient, *butoconazole*, we failed to map *5000 MG Butoconazole nitrate 20 MG/ML Prefilled Applicator* (*890780*) to *G01AF15*, because the dose form group of RxNorm drug, *Prefilled Applicator Product*, is not listed as compatible with the vaginal route, *V*, listed for this drug in ATC. Finally, some RxNorm drugs have no mapping to ATC because their ingredient is simply not present in ATC (e.g., *oregano allergenic extract 50 MG/ML Injectable Solution*).

Implementation

From a technical perspective, this investigation can be thought of as a data integration project. The datasets to be integrated include RxNorm, ATC, the ingredient mapping between RxNorm and ATC, and the mapping of both RxNorm dose form groups and ATC administration codes to administration routes. Semantic Web technologies are known to provide support for data integration. Here we converted all the datasets to the Resource Description Format (RDF triples) and loaded them into the triple store Virtuoso. The query language for RDF, SPARQL, also provides support for writing production rules (of the "if … then" type). We created production rules in order to generate the missing administration codes. We also created rules to perform the mapping of clinical drugs to ATC.

Evaluation of the mapping

Since the mapping of ingredients had been evaluated already and the mapping of the administration routes is straightforward, we are not concerned about false positives. The focus of our evaluation is a failure analysis, in which we reviewed the clinical drugs from RxNorm for which no mapping was found in ATC. The application of the mapping to characterizing a large prescription list provides additional opportunities for evaluation.

RxNorm clinical drugs with no ingredient mapping to ATC. One clinician (LT) reviewed the ingredients for which no mapping to ATC was found and classified them. A number of them revealed a known scope issue. A large number of allergenic extracts and other natural products from RxNorm are not in ATC. The remaining ingredients were classified into clinically important drugs and other drugs, with the objective of identifying drugs of clinical importance, which we had failed to map to ATC.

RxNorm clinical drugs with ingredient mapping, but no route mapping to ATC. This failure analysis is meant to identify cases where our mapping of administration routes would be deficient and prevent drugs from mapping between RxNorm and ATC. One clinician (LT) reviewed a subset of the most frequent unmapped RxNorm drugs and analyzed why no mapping to ATC was found, although the ingredient was present in both resources.

Results

Mapping ingredients between RxNorm and ATC

Of the 20,463 clinical drugs in RxNorm, 11,306 are single-ingredient, prescribable drugs are used in this investigation, corresponding to 2687 unique active moieties and 422 salt ingredients. There are 4516 5th-level codes in ATC, of which 3863 correspond to single active moieties, for a total of 3299 unique drug names.

We are able to map 1515 (56%) of the 2687 RxNorm active moieties to ATC, and 375 (89%) of its salt ingredients. Of note, most of the salt ingredients were mapped to active moieties rather than salt ingredients in ATC. These RxNorm ingredients were mapped to a total of 1962 5th-level ATC codes (1530 unique drug names), representing 51% of the 3863 5th-level ATC codes for single active moieties (and 46% of the drug names). In summary, about half of the RxNorm and ATC ingredients have a mapping to the other drug resource.

Mapping routes of administration between RxNorm and ATC

Harmonization of administration codes between RxNorm and ATC. The correspondence between RxNorm DFGs, ATC administration codes and the extended administration routes is shown in Table 3. Each of the 22 dose form groups from RxNorm and each of the 24 administration codes from ATC (including the ones we created) is mapped to one of the 15 administration routes (extended list). As a result, each dose form group from RxNorm can be associated with at least one administration code from ATC.

Table 3. Correspondence between RxNorm dose form groups and ATC administration codes through then extended list of administration routes derived from ATC.

RxNorm Dose Form Group	Route of administration	ATC Administration Code
Drug Implant Product	*implant*	implant s.c. implant
Inhalant Product	*inhalation*	Inhal Inhal.solution Inhal.powder Inh.aerosol Inhal.aerosol Instill.sol.
Nasal Product	*nasal*	N
Oral Product	*oral*	O Chewing gum oral aerosol
Ophthalmic Product	*ophthalmic*	lamella [ophthalmic]
Otic Product	*otic*	[otic]
Injectable Product	*parenteral*	P
Rectal Product	*rectal*	R
Sublingual Product	*sublingual/buccal*	SL
Buccal Product Dental Product Oral Cream Product Oral Foam Product Oral Gel Product Oral Ointment Product Oral Paste Product	*stomatologic*	[stomatologic]
Transdermal Product	*transdermal*	TD TD patch
Intraperitoneal Product Irrigation Product Mucosal Product Prefilled Applicator Product Shampoo Product Soap Product Topical Product	*topical*	intravesical ointment [topical]
Urethral Product	*urethral*	urethral
Vaginal Product	*vaginal*	V

Assignment of missing administration codes in ATC.

Of the 3863 5th-level codes in ATC for single active moieties, 2085 (54%) are missing an administration code. The distribution of the number of ATC codes for which administration codes were generated is listed in Table 4, by type of technique. Since the rules for ophthalmic, otic, stomatologic and rectal products were allowed to generate administration codes even when one had been asserted by ATC, the total number of ATC drugs for which administration codes were generated is slightly higher than the number of ATC codes with missing administration codes.

Table 4. Number of ATC codes for which administration codes were generated automatically, by type of technique.

Expression in ATC group label	#
Administration code inferred from ATC group	725
Administration code inferred from expressions found in the labels of ATC	232
Administration code inferred from drugs from the same ATC group	492
Oral admistration code inferred by default (digestive drugs)	23
Topical administration code inferred by default (remaining drugs)	643

Total	2115

Mapping RxNorm clinical drugs to ATC

Of the 11,306 clinical drugs from RxNorm, 7655 (68%) had an ingredient mapping to ATC, and 7170 (63%) had both an ingredient and an administration route mapping. In other words, a mapping between a clinical drug in RxNorm and a drug in ATC (at the 5th level) for a particular administration route was found for 63% of the clinical drugs in RxNorm.

These RxNorm clinical drugs mapped to 1693 unique ATC codes (86% of the 1962 ATC codes to which an ingredient mapping was found) and 1447 unique drug names (95% of the 1530 ATC drug names to which an ingredient mapping was found), corresponding to 44% of the 3863 ATC codes for single active moieties (and 44% of the drug names).

Evaluation

RxNorm clinical drugs with no ingredient mapping to ATC. Of the 1181 ingredients from RxNorm with no mapping to ATC, 818 (69%) correspond to allergenic extracts and other natural products (e.g., *kiwi fruit allergenic extract 50 MG/ML Injectable Solution*). A total of 29 ingredients were identified as clinically important, including hormones (e.g., *thyroid (USP) 97.5 MG Oral Tablet*), blood substitutes (e.g., *Dextran 1 150 MG/ML Injectable Solution*) and antihemorrhagics (especially coagulation factors). The remaining ingredients include uncommon drugs.

RxNorm clinical drugs with ingredient mapping, but no route mapping to ATC. In our exploration of the subset of the most frequent RxNorm drugs with no mapping to ATC, we found that when an ingredient does not map due to missing route, the routes in RxNorm and ATC generally do not match. Often, one particular route of administration is missing from RxNorm or ATC, but present in the other resource. For example, the RxNorm drug *5000 MG Butoconazole nitrate 20 MG/ML Prefilled Applicator* (890780) is a gynecologic product (with brand name *Gynazole-1*), but its dose form group, *Prefilled Applicator Product*, does not reflect a vaginal route. As a consequence, it does not match the vaginal form of *Butoconazole* in ATC (*G01AF15*). We also found cases where the missing routes of administration in ATC, which we assign through rules, are sometimes incomplete and could be refined. Overall, the largest route mapping failures happen when a topical route is assigned and the ingredient has more than one drug form.

Applications

We applied the mapping of RxNorm to ATC to the analysis of a large prescription set. This prescription set is coded to RxNorm and we use the mapping to ATC so that we can leverage the ATC classification of the drugs for characterizing the prescription list. We also compare the daily doses prescribed to the defined daily doses listed in ATC.

Coverage

Drugs from the SureScripts dataset used in this investigation had been coded to RxNorm. Using the RxNorm API, we updated obsolete codes to the current version of RxNorm and restricted drugs to the prescribable subset of RxNorm, in order to make them compatible with the clinical drugs present in pour mapping to ATC. Of the 100,195 RxNorm codes in the (updated) SureScripts dataset, 86,963 (87%) were found in the prescribable subset of RxNorm as single-ingredient drugs. The other drugs were either retired (e.g., *oxybutynin 5 MG Extended Release Tablet* (*312155*)), multi-ingredient drugs (e.g., *Amoxicillin 250 MG / Clavulanate 125 MG Oral Tablet* (*562251*)), or not included in the prescribable subset (e.g., *Potassium Chloride 10 MEQ Oral Tablet* (*312505*)).

Of the 86,963 RxNorm codes from this SureScript subset, 84,559 (97%) mapped to at least one code in ATC (through both the ingredient and the route). And of the 1692 distinct RxNorm clinical drugs found in the SureScript subset, 1611 (95%) were found in ATC.

Classification of drugs from a prescription list

Using the mapping to ATC, we classified the 84,559 prescriptions from the SureScripts set against the top-level categories in ATC. Since a clinical drug from RxNorm can map to several ATC codes (including across top-level groups), the total of ATC codes mapped to is slightly higher than the original number of prescription codes. However, this approach makes it easy to provide a profile of the drugs with respect to a widely used classification system. The distribution of SureScripts drugs by top-level ATC groups is shown in Table 5. The top categories are cardiovascular and nervous system drugs.

Table 5. Distribution of SureScripts drugs by top-level ATC groups.

ATC 1	Group label	Freq.
A	ALIMENTARY TRACT AND METABOLISM	9398
B	BLOOD AND BLOOD FORMING ORGANS	3189
C	CARDIOVASCULAR SYSTEM	21393
D	DERMATOLOGICALS	2949
G	GENITO URINARY SYSTEM AND SEX HORMONES	3337
H	SYSTEMIC HORMONAL PREPARATIONS, EXCL. SEX HORMONES AND INSULINS	4335
J	ANTIINFECTIVES FOR SYSTEMIC USE	7480
L	ANTINEOPLASTIC AND IMMUNOMODULATING AGENTS	1274
M	MUSCULO-SKELETAL SYSTEM	3476
N	NERVOUS SYSTEM	22841
P	ANTIPARASITIC PRODUCTS, INSECTICIDES AND REPELLENTS	357
R	RESPIRATORY SYSTEM	4680
S	SENSORY ORGANS	1919
V	VARIOUS	107
	Total	86735

Assessment of daily doses

Using the defined daily dose (DDD) provided by ATC for some ATC codes (and for a given route of administration), we analyzed the prescribed dose in SureScripts for these drugs. Of the 84,559 prescriptions from the SureScripts, 77,248 (91%) had a DDD (corresponding to 1317 unique RxNorm ingredients). Since the DDD is for an ingredient-route pair in ATC, we grouped the RxNorm drugs accordingly. After eliminating the cases where the prescribed dose was 0 (usually denoting “as needed”), we obtained 522 pairs and computed the median of the prescribed dose for each pair, which we compared to the DDD by normalizing the difference to the DDD. The distribution of the deviation from the DDD for the 522 pairs is shown in Figure 1. The main category (D: 134 ingredient-route pairs) correspond to drugs for which the median of the dose prescribed exactly corresponds to the DDD. Categories E-G correspond to drugs for which the dose prescribed exceeds the DDD, and categories B and C correspond to drugs for which the dose prescribed is lower than the DDD. The median daily dose prescribed is within +/-50% of the DDD for 60% of the 522 ingredient-route pairs in the prescription list analyzed.

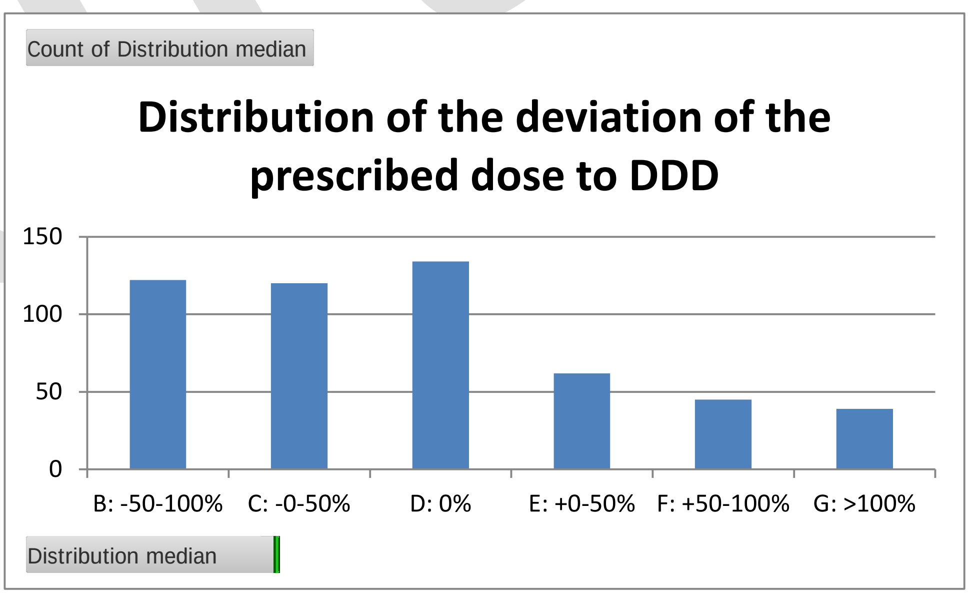

Figure 1. Distribution of the deviation from the DDD for the 522 ingredient-drug pairs.

Discussion and conclusion

Significance. Although the overall coverage for ingredient mapping is limited, we were able to demonstrate that prescription drugs in current use in the US are mapped reliably and in a considerable proportion. The comparison on DDD, and actual doses prescribed in clinical practice is encouraging. These two main findings in our work demonstrate that the ATC can be applied not only for its intended use of drug utilization evaluation, but also for the analysis of prescription datasets in the US.

Limitations. The largest limitation of our work is the relatively small size of the prescription dataset, in which the variation of drug ingredients is necessarily limited in scope, even more so in the case of drugs from emergency room patients only.

Future work. In addition to the exploration of larger and more diverse prescription datasets, the focus of future work is to refine the administration route assignment for missing routes in the ATC, allowing for more precise mapping.

Conclusion. Although the mapping of RxNorm ingredients to ATC is largely incomplete, the most frequently prescribed drugs in the prescription set we analyzed were covered.

Acknowledgments

This work was supported by the Intramural Research Program of the NIH, National Library of Medicine.

References

1. Volodina A, Bertsche T, Kostev K, Winkler V, Haefeli WE, Becher H. Drug utilization patterns and reported health status in ethnic German migrants (Aussiedler) in Germany: a cross-sectional study. BMC public health 2011;11:509 http://www.ncbi.nlm.nih.gov/pubmed/21711531.
2. Rodriguez-Ganen O, Asbun-Bojalil J. [Surveillance of antimicrobial drug use in Mexican hospitals: current situation and practical guidelines for its implementation]. Revista panamericana de salud publica = Pan American journal of public health 2012;32(5):381-6 http://www.ncbi.nlm.nih.gov/pubmed/23338696.
3. Haug JB, Myhr R, Reikvam A. Pharmacy sales data versus ward stock accounting for the surveillance of broad-spectrum antibiotic use in hospitals. BMC medical research methodology 2011;11:166 http://www.ncbi.nlm.nih.gov/pubmed/22166018.
4. Guenette L, Moisan J. Elderly people's knowledge of the purpose of their medicines. The American journal of geriatric pharmacotherapy 2011;9(1):49-57 http://www.ncbi.nlm.nih.gov/pubmed/21459308.
5. Avillach P, Dufour JC, Diallo G, Salvo F, Joubert M, Thiessard F, et al. Design and validation of an automated method to detect known adverse drug reactions in MEDLINE: a contribution from the EU-ADR project. Journal of the American Medical Informatics Association : JAMIA 2012 http://www.ncbi.nlm.nih.gov/pubmed/23195749.
6. Chen L, Zeng WM, Cai YD, Feng KY, Chou KC. Predicting Anatomical Therapeutic Chemical (ATC) classification of drugs by integrating chemical-chemical interactions and similarities. PloS one 2012;7(4):e35254 http://www.ncbi.nlm.nih.gov/pubmed/22514724.
7. Gurulingappa H, Kolarik C, Hofmann-Apitius M, Fluck J. Concept-based semi-automatic classification of drugs. Journal of chemical information and modeling 2009;49(8):1986-92 http://www.ncbi.nlm.nih.gov/pubmed/19663460.
8. Winnenburg R, Bodenreider O. Exploring pharmacoepidemiologic groupings of drugs from a clinical perspective. Stud Health Technol Inform (Proc Medinfo) 2013:(accepted)
9. Winnenburg R, Bodenreider O. Mapping drug entities between the European and American standards, ATC and RxNorm. Poster Proceedings of the Eighth International Conference on Data Integration in the Life Sciences (DILS 2012) 2012:22